A NOTE ON ADDITIVITY

BU-101-M

Walter T. Federer*

January, 1959

The linear model for a two-way classification with one observation per cell may be expressed as

$$X_{ij}=\mu+\alpha_i+\beta_j+\epsilon_{ij}=\bar{x}+(\bar{x}_{i.}-\bar{x})+(\bar{x}_{.j}-\bar{x})+e_{ij}$$

where $\bar{x}_{i.}$, $\bar{x}_{.j}$ and $\bar{x}$ are the ith row mean, the jth column mean and the overall mean, respectively. Denote this as model s.

In a chi-square contingency table the linear model for the computed or "expected" value is obtained as

$$\frac{X_{i.}X_{.j}}{X_{..}}=\frac{\bar{x}_{i.}\bar{x}_{.j}}{x}=\bar{x}_{i.}+\bar{x}_{.j}-\bar{x}+(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})/\bar{x}$$

$$=\bar{x}+(\bar{x}_{i.}-\bar{x})+(\bar{x}_{.j}-\bar{x})+\frac{(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{x}} \quad .$$

Denote the above plus an error term d_{ij} as model c.

The interaction or residual sum of squares under model s is computed as

$$\sum_{i}\sum_{j} e_{ij}^2=\sum_{i}\sum_{j}(X_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x})^2=\sum_{i}\sum_{j}(X_{ij}-X'_{ij})^2 \quad ;$$

under model c the interaction sum of squares is computed as

$$\sum_{i}\sum_{j} d_{ij}^2=\sum_{i}\sum_{j}\left(X_{ij}-\frac{\bar{x}_{i.}-\bar{x}_{.j}}{\bar{x}}\right)^2=\sum\sum(X_{ij}-X''_{ij})^2 \quad .$$

The difference in the interaction sums of squares under model s and under the more restricted model, model c, is

$$\sum_{i}\sum_{j}(X_{ij}-X'_{ij})^2-\sum_{i}\sum_{j}(X_{ij}-X''_{ij})^2$$

$$=\sum_{i}\sum_{j}\left\{2X_{ij}\frac{(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{x}}-\left(\frac{(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{x}}\right)^2\right\}$$

$$=\sum_{i}\sum_{j}\frac{(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{x}}\left\{2X_{ij}-\frac{(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{x}}\right\} \quad .$$

---

*Biometrics Unit, Plant Breeding Department, Cornell University

... next to the last form appears to be the better one for computational purposes; the last form does not appear to reduce further. The above sum of squares is not always positive.

Now let's compare this difference in sums of squares with Tukey's one degree of freedom sum of squares for non-additivity (TNA), which is computed from the formula

$$\frac{\left\{\sum_{ij}\sum X_{ij}(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})\right\}^2}{\sum_i(\bar{x}_{i.}-\bar{x})^2\sum_j(\bar{x}_{.j}-\bar{x})^2}.$$

Also, we shall denote the former sum of squares as a row-column-association (RCA) and allocate one degree of freedom to it since it measures the association of row deviations and column deviations. To make comparisons of TNA and RCA sums of squares, the five numerical examples given by Harter and Lum were used.

The first example is

	Column					
Row	1	2	3	$X_{i.}$	$\bar{x}_{i.}$	$\bar{x}_{i.}-\bar{x}$
1	2	4	12	18	6	-3
2	4	8	24	36	12	3
$X_{.j}$	6	12	36	54	---	0
$\bar{x}_{.j}$	3	6	18	--	9	--
$\bar{x}_{.j}-\bar{x}$	-6	-3	9	0	---	---

The sums of squares in the analysis of variance are:

Source of variation	d.f.		Sum of squares	
Row	1		54	
Column	2		252	
Interaction	2		28	
TNA		1		28
Residual		1		0
Row-column Association (RCA)		1		28

The second example is

Row	Column 1	Column 2	Column 3	$X_{i.}$	$\bar{x}_{i.}$	$\bar{x}_{i.}-\bar{x}$
1	21	30	54	105	35	-23
2	75	78	90	243	81	23
$X_{.j}$	96	108	144	348	--	0
$\bar{x}_{.j}$	48	54	72	--	58	--
$\bar{x}_{.j}-\bar{x}$	-10	-4	14	0	--	--

The sums of squares in the analysis of variance are:

Source of variation	d.f.		Sum of squares	
Row	1		3174	
Column	2		624	
Interaction	2		84	
TNA		1		83.3077
Residual		1		0.6923
RCA		1		82.7016

where $\frac{2(23)|-228|}{58} - \frac{23^2(624)}{58^2} = 82.7016$

The third example is:

Row	Column 1	Column 2	Column 3	$X_{i.}$	$\bar{x}_{i.}$	$\bar{x}_{i.}-\bar{x}$
1	36	81	144	261	87	36
2	0	9	36	45	15	-36
$X_{.j}$	36	90	180	306	--	0
$\bar{x}_{.j}$	18	45	90	--	51	--
$\bar{x}_{.j}-\bar{x}$	-33	-6	39	0	--	--

The sums of squares in the analysis of variance are:

Source of variation	d.f.		Sum of squares	
Row	1		7776	
Column	2		5292	
Interaction	2		1296	
TNA		1		1269.55
Residual		1		26.45
RCA		1		1022.45

The fourth example is:

Row	Column 1	Column 2	Column 3	$X_{i.}$	$\bar{x}_{i.}$	$\bar{x}_{i.}-\bar{x}$
1	60	90	180	330	110	-37
2	72	120	360	552	184	37
$X_{.j}$	132	210	540	882	--	0
$\bar{x}_{.j}$	66	105	270	--	147	--
$\bar{x}_{.j}-\bar{x}$	-81	-42	123	0	--	--

The sums of squares in the analysis of variance are:

Source of variation	d.f.		Sum of squares	
Row	1		8214	
Column	2		46908	
Interaction	2		8508	
TNA		1		8449.06
Residual		1		58.94
RCA		1		7049.94

The last example is:

Row	Column 1	Column 2	Column 3	$X_{i.}$	$\bar{x}_{i.}$	$\bar{x}_{i.}-\bar{x}$
1	10	11	3	24	8	-1
2	6	9	15	30	10	1
$X_{.j}$	16	20	18	54	---	0
$\bar{x}_{.j}$	8	10	9	---	9	---
$\bar{x}_{.j}-\bar{x}$	-1	1	0	0	---	---

The sums of squares in the analysis of variance are:

Source of variation	d.f.		Sum of squares	
Row	1		6	
Column	2		4	
Interaction	2		76	
TNA		1		1
Residual		1		75
RCA		1		32/81

According to Harter and Lum if X^p should be used instead of X and if $p < 1$ then the sum of squares for RCA as computed above will always be positive. From the numerical examples numbers 1,3,4, and 5 it would appear that the sum of squares for RCA is less than or equal to the sum of squares for TNA, i.e.,

$$\frac{[\Sigma\Sigma X_{ij}(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})]^2}{\Sigma(\bar{x}_{i.}-\bar{x})^2\Sigma(\bar{x}_{.j}-\bar{x})^2} \geq \frac{2\Sigma\Sigma X_{ij}(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{\bar{x}}}$$

$$-\Sigma\Sigma\left[\frac{(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{\bar{x}}}\right]^2 \qquad (\text{for } p < 1) \quad .$$

There does, however, appear to be a high relationship between the sums of squares for TNA and RCA.

When $p \geq 1$ the sum of squares,

$$\Sigma\Sigma X_{ij}(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})$$

is negative. For example 2, the quantity

$$2\left|\frac{\Sigma\Sigma X_{ij}(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{\bar{x}}}\right| - \Sigma\Sigma\left(\frac{(\bar{x}_{i.}-\bar{x})(\bar{x}_{.j}-\bar{x})}{\bar{\bar{x}}}\right)^2 = 82.7016$$

did agree well with the sum of squares for TNA = 83.3077. Whether this is coincidence or fact could be determined from further empirical investigations of this sort, or from algebraic manipulation of the above formulae.

The important subject of additivity of data requires additional investigation before the needs of experimenters can be adequately satisfied. Professor John W. Tukey has made notable advances and others have made contributions toward the solution of the statistical problems associated with the non-additivity problem (see references at end of paper).

The material in this paper represents a different approach than used by other authors. The problem of viewing non-additivity in another light is presented for its novelty and contribution to the entire problem rather than for potential usefulness in applied statistics.

References

1. Anscombe, F. J., and Tukey, J. W., The criticism of transformations. J. Amer. Stat. Assoc. 50:566, 1955 (abstract).

2. Federer, W. T., Experimental design, pp. 49-51 and 207-209, Macmillan, N. Y., 1955.

3. Harter, H. L., and Lum, Mary, D., A note on Tukey's one degree of freedom for non-additivity. Paper presented at Amer. Stat. Assoc. meetings, Atlantic City, N. J., 9/10-13/57.

4. Moore, P. G., and Tukey, J. W., Answer to Query 112. Biometrics 10:562-568, 1954.

5. Tukey, J. W., One degree of freedom for non-additivity. Biometrics 5:232-242, 1949.

6. Tukey, J. W., Answer to Query 113. Biometrics 11:111-113, 1955.

7. Tukey, J. W., On the comparative anatomy of transformations. Annals of Math. Stat. 28:602-632, 1957.